MARKERLESS ACTIVE TRUNK SHAPE MODELLING FOR MOTION TOLERANT REMOTE RESPIRATORY ASSESSMENT

Vahid Soleimani[*]*, Majid Mirmehdi*[*]*, Dima Damen*[*] *and James Dodd*[†]

[*]Visual Information Laboratory, Faculty of Engineering, University of Bristol, BS8 1UB, UK
vahid.soleimani@bristol.ac.uk
[†]Academic Respiratory Unit, Southmead Hospital, Bristol, BS10 5NB, UK

ABSTRACT

We present a vision-based trunk-motion tolerant approach which estimates lung volume–time data remotely in forced vital capacity (FVC) and slow vital capacity (SVC) spirometry tests. After temporal modelling of trunk shape, generated using two opposing Kinects in a sequence, the chest-surface respiratory pattern is computed by performing principal component analysis on temporal geometrical features extracted from the chest and posterior shapes. We evaluate our method on a publicly available dataset of 35 subjects (300 sequences) and compare against the state-of-the-art. By filtering complex trunk motions, our proposed method calibrates the entire volume–time data using only the *tidal volume* scaling factor which reduces the state-of-the-art average normalised L_2 error from 0.136 to 0.05.

Index Terms— trunk shape modelling, noncontact vision-based respiratory assessment, chest shape, spirometry.

1. INTRODUCTION

Lung function diseases are among the leading causes of death worldwide. In 2015, Chronic Obstructive Pulmonary Disease (COPD) and Asthma together affected $\sim$532 million people worldwide, of whom $\sim$3.6 million died [1].

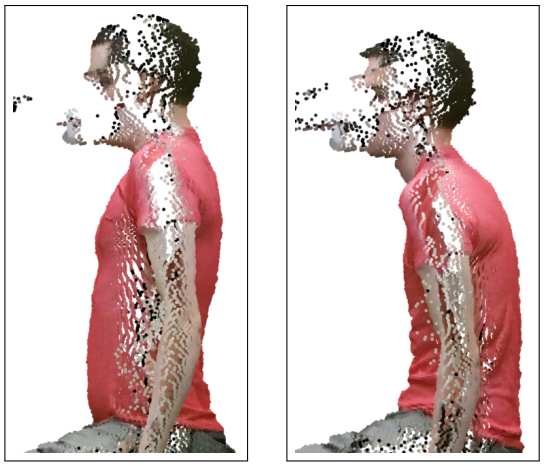
Spirometry is a clinically approved method for diagnosis and assessment of lung function diseases. A routine spirometry test is performed in an upright sitting posture, in which patients breathe into a contact-based pneumotach in a particular pattern, depending on the intended lung function test. Two primary clinical spirometry tests are forced vital capacity (FVC) and slow vital capacity (SVC). Both tests start with a few cycles of normal breathing, called *tidal volume*, followed by a maximal *inhalation–exhalation*, called *main effort*. The *main effort inhalation–exhalation* is performed at the same speed of normal breathing in the SVC test, whereas it is performed as fast and forcefully as possible in the FVC test. Using respiratory volume–time data obtained from a pneumotach, several pulmonary function testing (PFT) measures, e.g., *FVC*, *FEV1*, *PEF*, $FEF_{25-75\%}$, *TV*, *ERV*, are computed and used for the diagnosis of pulmonary function diseases [2].

Although spirometry is a reliable clinical approach, it has several drawbacks which reduce its suitability and usability. As a contact-based device, a pneumotach is intrusive and difficult for anyone to perform with it, especially frail elderly, children and cognitively impaired patients. To measure accurately enough, a pneumotach must be manually recalibrated before each session. Further, spirometry is rather expensive given the cost of the required consumables (mouthpiece and nose-clip) and the price of the pneumotach itself.

Recently, we introduced a vision-based pulmonary function testing approach [3–5] which addresses spirometry drawbacks by remotely computing 11 FVC and SVC measures, e.g., *FVC*, *FEV1*, *PEF*, $FEF_{25-75\%}$, *TV*, *ERV*. These PFT measures are computed from calibrated depth-based volume–time data obtained by estimating the chest volume variation using depth measurements obtained from a single [3, 4] or two depth sensors [5]. The calibration process linearly scales the estimated chest volume to the real lung volume, and is required to compute correct PFT measures. The calibration process is performed using subject-specific (intra-subject) scaling factors, learnt in a training phase from both depth-based and spirometer volume–time data. The medical achievements of our remote approaches were reported in [6, 7]. Apart from our works [3–7], we know of only [8,9] which also performed remote respiratory assessment, rather than just breathing rate estimation or respiration monitoring [10–15]. In [8, 9], Ostadabbas et al. detected airway obstruction as mild, moderate and severe, and computed only *FEV1* in [8]. To acquire only chest-surface respiratory motion, they heavily restrained trunk motions in both works.

As a natural reaction of the human respiratory system, subjects inevitably move their trunk during PFT, especially at the *main effort inhalation–exhalation* stage. Since constraining such reactive motions would affect the lung function measures, PFT must be performed as in routine spirometry, without restraint. However, when a single depth sensor is used, decoupling trunk motion and chest-surface respiratory mo-

The authors would like to thank the University of Bristol Alumni Foundation for funding this research, and SPHERE for providing the opportunity of collaboration between engineering and clinical researchers.



(a) deep *inhalation* (b) deep forced *exhalation*

Fig. 1. Natural reaction of subject's trunk to (a) deep *inhalation*, and (b) deep forced *exhalation*, in *main effort* breathing.

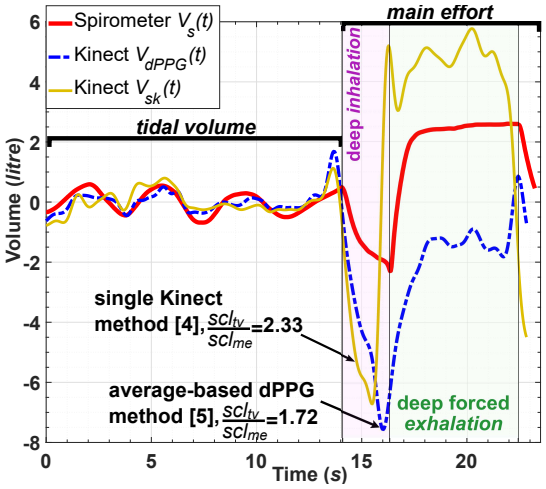


Fig. 2. The single Kinect [4] and dPPG [5] volume–time data and their *tidal volume* to *main effort* scaling factors ratio.

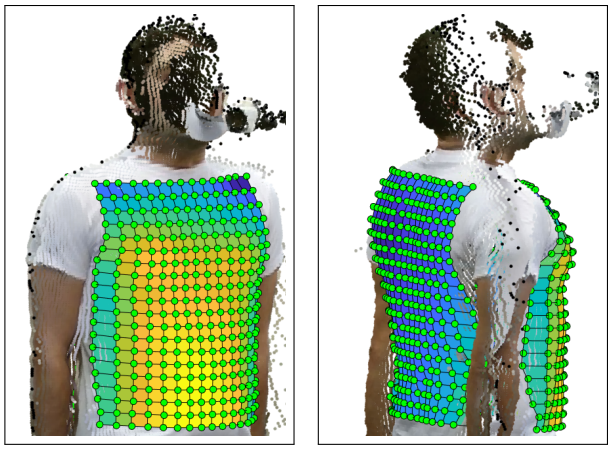
tion would be potentially impossible, and our single Kinect method [3, 4] was therefore influenced in three ways. First, PFT measures were directly affected as they were computed from volume–time data corrupted by trunk motion. Second, while the *calibration* scaling factors were supposed to remain unchanged for a subject in all PFT performances (trials), trunk motion caused inconsistent training scaling factors, consequently decreasing the accuracy of the computed PFT measures. Third, different patterns of trunk motion during *tidal volume* and *main effort* breathing, required their data to be calibrated individually using distinct scaling factors, whereas in theory the whole volume–time data (*tidal volume* + *main effort*) must be calibrated using a single scaling factor.

In our most recent work [5], we presented a depth-based whole body photoplethysmography (dPPG) approach which filtered the trunk motion by constructing a 3-D model of a subject's trunk and then subtracting the average depth of the chest-wall from the average depth of the posterior-wall per frame. While the dPPG method showed significant advancement in the accuracy of the computed PFT measures and also in the consistency of the scaling factors, there is still room for further improvements. First, the dPPG method cannot filter complex trunk motion patterns, particularly at the *main effort inhalation–exhalation*, due to using only the average depth of the chest and posterior walls. Second, different patterns of trunk motion in *tidal volume* and *main effort* breathing, requires individual calibration of their volume–time data rather than calibrating the whole volume–time data using a single scaling factor. For example, Fig. 1 (a) and (b) show the body's natural reaction to the *main effort* deep *inhalation* and deep forced *exhalation* during a routine spirometry FVC test, respectively. Correspondingly, Fig. 2 presents the volume–time data of the single Kinect $V_{sk}(t)$ [4] and dPPG $V_{dPPG}(t)$ [5] methods for this PFT test which are entirely calibrated using only the *tidal volume* scaling factor. As seen, both the single Kinect and dPPG methods fail to provide an accurate enough volume–time data, though the dPPG performed better. Also, as the ratios of the *tidal volume* to *main effort* scaling factors in each case, i.e., 2.33 and 1.72, denote, both methods obtained considerably different scaling factors. This can be perceived by comparing the scale of depth-based *main effort* to the spirometer's *main effort* in Fig. 2.

In this work, we propose a novel trunk shape modelling approach to extract the chest-surface respiratory pattern by temporal modelling and tracking of trunk shape, generated using the depth measurements acquired by two opposing depth sensors. Instead of filtering the trunk motion by subtracting the average depth of the chest from the average depth of the posterior in [5], here we extract the chest-surface respiratory pattern by performing a principal component analysis (PCA) on temporal 3-D geometrical features extracted from the chest and posterior shape models in $\mathbb{R}^3$ space. To present the real volume of exchanged air, the respiratory volume–time data is then calibrated using scaling factors learnt in a training phase. We validate our method on a dataset of 35 subjects (300 sequences) introduced in [5], by computing the normalised L_2 error, dynamic time warping and Fréchet distances, as well as correlation of determination between the depth-based and spirometer volume–time data.

2. PROPOSED METHODOLOGY

We acquire the subject's 3-D body data using the open source data acquisition and registration pipeline introduced in [16]. This pipeline synchronises and calibrates two opposing Kinects by which we create an almost complete 3-D model of a subject with a consistent frame rate (30*fps*) during PFT performance (Figs. 1 and 3). Using skeletal joint



(a) Front view (b) Side view

Fig. 3. A subject's chest and posterior shape models with their data points presented as small green spheres on them.

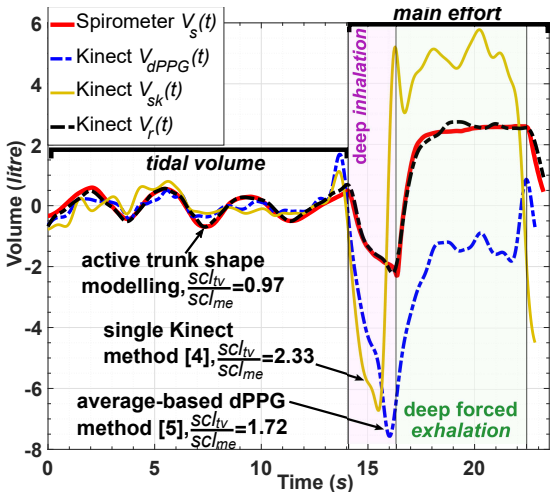


Fig. 4. Comparing volume–time data and scaling factors ratio of the proposed method to the single Kinect [4] and dPPG [5].

data, a 3-D mask is automatically generated and used to filter inessential body parts. Fig. 3 shows a sample of a trunk shape model laid over the subject's reconstructed body.

Although the dPPG method [5] was able to decouple the trunk motion to a notable extent, it failed to correctly filter the complex trunk motion patterns due to, (i) using the depth as the only feature of the chest and posterior walls, i.e., z-coordinate of their data points in $\mathbb{R}^3$ space, and (ii) eliminating useful potential features by filtering the chest and posterior data points to single average depth values. We address these issues by extracting a set of temporal geometrical features over the entire chest and posterior walls within a PFT sequence. Using these features, partial volumes are uniformly computed from the full trunk, and tracked per frame. However, as a markerless trunk reconstruction approach, we are faced with two main challenges for extracting the proposed features. First, the number and location of data points in either of the chest and the posterior regions may vary from one frame to another. Second, since each Kinect captures either the chest or the posterior, these two regions are different in the number of the data points and their locations. Thus, it is quite unlikely that there would be a corresponding co-located posterior data point for a chest data point in the xy-plane.

We address these issues by (a) defining a fixed size 2-D region $\boldsymbol{\mathcal{R}}_{xy}(t)$ which covers the chest in the xy-plane at time t through the sequence, and generating a fixed number of points uniformly distributed in it, and (b) defining temporal interpolant functions $\boldsymbol{\mathcal{I}}^t_{ch}$ and $\boldsymbol{\mathcal{I}}^t_{po}$ to compute corresponding chest and posterior data points in $\mathbb{R}^3$ space which are located at the same position in the xy-plane. Fig. 3 shows the data points on the chest and posterior shapes as small green spheres. The interpolant functions are generated by applying a Delaunay triangulation on the chest and the posterior original data points [17]. Using $\boldsymbol{\mathcal{I}}^t_{ch}$ and $\boldsymbol{\mathcal{I}}^t_{po}$, partial volumes are computed over the trunk as

$$\mathbf{v}_{ij}(t) = \left[\boldsymbol{\mathcal{I}}^t_{po}\Big(\mathbf{p}_{ij}(t)\Big) - \boldsymbol{\mathcal{I}}^t_{ch}\Big(\mathbf{p}_{ij}(t)\Big)\right], \forall \mathbf{p}_{ij}(t) \in \boldsymbol{\mathcal{R}}_{xy}(t). \quad (1)$$

Since the extent of body motion along the trunk during PFT varies, i.e., from the minimum at the hips to its maximum at the shoulders, we accumulate $\mathbf{v}_{ij}(t)$ along the longitudinal direction to further reduce their sensitivity to the trunk motion. Thus, the final feature matrix for the whole sequence is created by extracting these partial volumes and their location in the xy-plane for each frame of the sequence as

$$\mathbf{F} = \Big[\Big\langle \boldsymbol{\mathcal{R}}_{xy}(t), \textstyle\sum_j \mathbf{v}_{ij}(t) \Big\rangle\Big]_{t=0}^{t=\tau}. \quad (2)$$

Finally, PCA is performed on the feature matrix $\mathbf{F}$ by which the respiratory volume–time data $V_r(t)$ is provided by the first principal component. $V_r(t)$ is then smoothed using a 4th order Butterworth low-pass filter to eliminate high frequency noise. Given a wide range of respiratory frequencies for adults and elderly at $0.2-0.6$ Hz [18], the cut-off frequency was chosen as 1.2 Hz to ensure preserving the respiratory data. Fig. 4 presents $V_r(t)$ and its comparison to the volume–time data obtained by the spirometer $V_s(t)$, single Kinect $V_{sk}(t)$ [4] and dPPG $V_{dPPG}(t)$ [5] methods. As seen, while $V_{sk}(t)$ and $V_{dPPG}(t)$ have been significantly affected by the trunk motion, especially at the deep forced *inhalation–exhalation* stage, $V_r(t)$ is much more accurate.

To present the real volume of exchanged air, $V_r(t)$ is calibrated by linearly scaling the y-axis using scaling factors learnt per subject in a *training* phase from spirometer and depth-based volume–time data. In the *training* phase, corresponding data samples of $V_s(t)$ and $V_r(t)$ are identified by sampling $V_s(t)$ at 30 Hz, and detrending $V_s(t)$ and $V_r(t)$, and temporally aligning them. Next, the *tidal volume* and *main effort* scaling factors are separately computed by solving linear

Table 1. Evaluation results of 155 FVC and 145 SVC trials, calibrated using *tidal volume* and *main effort* scaling factors

	Metric	Single Kinect [4]	dPPG Method [5]	Proposed Method
FVC test	$\mathbf{NL_2}$	3.118 ± 13.88	0.059 ± 0.097	$\mathbf{0.013 \pm 0.016}$
	FRD	55.14 ± 102.93	12.38 ± 9.13	$\mathbf{6.36 \pm 3.40}$
	DTW	808.9 ± 1764.8	113.2 ± 104.8	$\mathbf{54.30 \pm 47.90}$
	$\mathbf{R^2}$	0.60 ± 0.32	0.84 ± 0.19	$\mathbf{0.96 \pm 0.04}$
SVC test	$\mathbf{NL_2}$	0.763 ± 3.311	0.032 ± 0.042	$\mathbf{0.015 \pm 0.016}$
	FRD	36.71 ± 56.97	12.06 ± 5.78	$\mathbf{8.36 \pm 4.26}$
	DTW	586.3 ± 1069.0	128.8 ± 96.02	$\mathbf{67.62 \pm 42.89}$
	$\mathbf{R^2}$	0.75 ± 0.23	0.92 ± 0.07	$\mathbf{0.96 \pm 0.03}$

Table 2. Evaluation results of 155 FVC and 145 SVC trials, calibrated using only *tidal volume* scaling factor

	Metric	Single Kinect [4]	dPPG Method [5]	Proposed Method
FVC test	$\mathbf{NL_2}$	0.679 ± 1.243	0.181 ± 0.309	$\mathbf{0.051 \pm 0.079}$
	FRD	39.31 ± 35.51	21.08 ± 17.89	$\mathbf{11.95 \pm 8.08}$
	DTW	541.4 ± 566.1	248.3 ± 246.7	$\mathbf{136.2 \pm 125.8}$
	$\mathbf{R^2}$	0.59 ± 0.32	0.78 ± 0.28	$\mathbf{0.93 \pm 0.06}$
SVC test	$\mathbf{NL_2}$	0.229 ± 0.327	0.092 ± 0.137	$\mathbf{0.049 \pm 0.047}$
	FRD	31.15 ± 20.64	19.72 ± 13.35	$\mathbf{15.49 \pm 7.34}$
	DTW	510.2 ± 470.9	279.6 ± 250.6	$\mathbf{204.1 \pm 132.6}$
	$\mathbf{R^2}$	0.73 ± 0.23	0.91 ± 0.06	0.90 ± 0.07

least square equations for $V_s(t)$ and $V_r(t)$ using an overdetermined system. This process is repeated for every pair of training spirometer and depth-based volume–time data which provides several pairs of training scaling factors.

For a test sequence, $V_r(t)$ is first computed using the proposed trunk shape modelling approach, and then calibrated by an average of all the training scaling factors. The training and testing process was carried out using leave-one-out cross-validation due to the limited number of trials per subject.

3. EXPERIMENTAL RESULTS

The evaluation was performed on a publicly available dataset (http://doi.org/ckrh) of 27 males and 8 females comprising 300 PFT sequences, i.e., 155 FVC and 145 SVC, recently introduced by [5]. Our method is able to process and retrieve the volume–time data of 2 sequences which were not used in [5] due to their complex trunk motion.

We evaluate our method in the signal level by computing (i) $\mathrm{NL}_2 = \sum_{t=0}^{\tau} \big(V_s(t) - V_r(t)\big)^2 / \big(\tau * (V_s^{max} - V_s^{min})\big)$, (ii) Fréchet distance (FRD) [19], (iii) dynamic time warping distance (DTW), and (iv) correlation of determination (R^2), between the depth-based and spirometer volume–time data. Tables 1 and 2 report the mean $\pm$ SD of these metrics across the 300 PFT sequences for two sets of differently calibrated volume–time data. In the first set, the data was calibrated using individual *tidal volume* and *main effort* scaling factors, whereas in the next set it was calibrated using only the *tidal volume* scaling factor. Our method outperforms the single Kinect [4] and dPPG [5] methods across all of the evaluation metrics in both calibration methods, except for R^2 of SVC test in Table 2 where our method achieves almost the same R^2 as dPPG. For example, the average NL_2 error across 300 sequences has been reduced from 1.94 and 0.045 by the single Kinect [4] and dPPG [5] methods to 0.014 by our method for individual calibration of *tidal volume* and *main effort*. It is clear that the more accurate the volume–time is, the more accurate the extracted PFT measures would be.

Comparing Tables 1 and 2 verifies that calibrating the entire volume–time data using only the *tidal volume* scaling factor achieves lower accuracy than the individual calibration of *tidal volume* and *main effort*. However, as a significant achievement, the difference in this accuracy is notably lower in our approach than the single Kinect [4] and dPPG [5] methods, e.g. while FVC average NL_2 error increases by 0.038 in our method, it increases by 0.122 in the dPPG. Calibrating the entire volume–time data using only the *tidal volume* scaling factor eliminates the threshold-based analysis required for separate calibration of *tidal volume* and *main effort*, which not only allows our approach to perform on-line, but also removes the calibration error due to incorrect keypoints computation.

Table 3. Comparing $\mathrm{R_{TM}}$ for 155 FVC and 145 SVC trials

Metric	Single Kinect [4]	dPPG Method [5]	Proposed Method
FVC $\mathbf{R_{TM}}$	1.11 ± 0.61	1.27 ± 0.41	$\mathbf{1.07 \pm 0.17}$
SVC $\mathbf{R_{TM}}$	1.11 ± 0.56	1.30 ± 0.41	$\mathbf{1.10 \pm 0.14}$

To further evaluate our method in filtering different patterns of trunk motion in *tidal volume* and *main effort*, we computed the mean $\pm$ SD of *tidal volume* to *main effort* scaling factors ratio ($\mathrm{R_{TM}}$) across all our sequences (see Table 3). Our method obtains the closest ratio to 1 with the lowest SD, achieving very close *tidal volume* and *main effort* scaling factors. This implies that we can calibrate the entire volume–time data using only the *tidal volume* scaling factor which is more reliable and less affected by the trunk motion.

4. CONCLUSION

We proposed a trunk-motion tolerant approach for estimating respiratory volume–time data within forced and slow vital capacity tests. Using temporal geometrical features extracted from the chest and posterior shapes, a feature matrix is created for the whole PFT sequence from which the respiratory volume–time data is computed by performing a principal component analysis. Evaluation on a dataset of 300 PFT sequences verifies that our method achieves more accurate volume–time data and also closer *tidal volume* and *main effort* scaling factors, compared to the state-of-the-art.

5. REFERENCES

[1] GBD, "Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015," *The Lancet Respiratory Medicine*, vol. 5, no. 9, pp. 691–706, 2017.

[2] Martin R Miller, JATS Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, et al., "Standardisation of spirometry," *The European Respiratory Journal*, vol. 26, no. 2, pp. 319–338, 2005.

[3] Vahid Soleimani, Majid Mirmehdi, Dima Damen, Sion Hannuna, Massimo Camplani, Jason Viner, and James Dodd, "Remote pulmonary function testing using a depth sensor," in *BioCAS*, 2015.

[4] Vahid Soleimani, Majid Mirmehdi, Dima Damen, James Dodd, Sion Hannuna, Charles Sharp, Massimo Camplani, and Jason Viner, "Remote, depth-based lung function assessment," *IEEE Transaction on Biomedical Engineering*, vol. 64, no. 8, pp. 1943–1958, 2017.

[5] Vahid Soleimani, Majid Mirmehdi, Dima Damen, Massimo Camplani, Sion Hannuna, James Dodd, and Charles Sharp, "Depth-based whole body photoplethysmography in remote pulmonary function testing," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1421–1431, 2018.

[6] Charles Sharp, Vahid Soleimani, Sion Hannuna, Massimo Camplani, Dima Damen, Jason Viner, Majid Mirmehdi, and James W Dodd, "Remote pulmonary function testing–computer gaming in the respiratory world," *Thorax*, vol. 70, no. Suppl 3, 2015.

[7] Charles Sharp, Vahid Soleimani, Sion Hannuna, Massimo Camplani, Dima Damen, Jason Viner, Majid Mirmehdi, and James W Dodd, "Towards respiratory assessment using depth measurements from a time-of-flight sensor," *Frontiers in Physiology*, vol. 8, no. 65, 2017.

[8] Sarah Ostadabbas, Christoph Bulach, David N Ku, Larry J Anderson, and Maysam Ghovanloo, "A passive quantitative measurement of airway resistance using depth data," in *EMBC*, 2014, pp. 5743–5747.

[9] Sarah Ostadabbas, Nordine Sebkhi, Mingxi Zhang, Salman Rahim, Larry J Anderson, Frances Eun-Hyung Lee, and Maysam Ghovanloo, "A vision-based respiration monitoring system for passive airway resistance estimation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1904–1913, 2016.

[10] Flavia Benetazzo, Alessandro Freddi, Andrea Monteriù, and Sauro Longhi, "Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results," *Healthcare Technology Letters*, vol. 1, no. 3, pp. 81–86, 2014.

[11] Dangdang Shao, Yuting Yang, Chenbin Liu, Francis Tsow, Hui Yu, and Nongjian Tao, "Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2760–2767, 2014.

[12] Rik Janssen, Wenjin Wang, Andreia Moço, and Gerard de Haan, "Video-based respiration monitoring with automatic region of interest detection," *Phys. Measurement*, vol. 37, no. 1, 2016.

[13] Kuan-Yi Lin, Duan-Yu Chen, and Wen-Jiin Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3263–3271, 2016.

[14] Bersain A Reyes, Natasa Reljin, Youngsun Kong, Yunyoung Nam, and Ki H Chon, "Tidal volume and instantaneous respiration rate estimation using a smartphone camera," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 764–777, 2017.

[15] Michael H Li, Azadeh Yadollahi, and Babak Taati, "Noncontact vision-based cardiopulmonary monitoring in different sleeping positions," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1367–1375, 2017.

[16] Vahid Soleimani, Majid Mirmehdi, Dima Damen, Sion Hannuna, and Massimo Camplani, "3d data acquisition and registration using two opposing kinects," in *3D Vision*, 2016, pp. 128–137.

[17] Isaac Amidror, "Scattered data interpolation methods for electronic imaging systems: a survey," *Journal of electronic imaging*, vol. 11, no. 2, pp. 157–176, 2002.

[18] Michelle A Cretikos, Rinaldo Bellomo, Ken Hillman, Jack Chen, Simon Finfer, and Arthas Flabouris, "Respiratory rate: the neglected vital sign," *Medical Journal of Australia*, vol. 188, no. 11, pp. 657–659, 2008.

[19] Boris Aronov, Sariel Har-Peled, Christian Knauer, Yusu Wang, and Carola Wenk, "Fréchet distance for curves, revisited," in *ESA*. Springer, 2006, vol. 4168, pp. 52–63.